BOREAL ENVIRONMENT RESEARCH 21: 363–371
ISSN 1239-6095 (print) ISSN 1797-2469 (online)

Helsinki 23 May 2016

Identification of boundaries between size classes of atmospheric aerosol particles

Hannes Tammet

Institute of Physics, University of Tartu, Ülikooli 18, EE-50090 Tartu, Estonia

Received 28 Oct. 2015, final version received 29 Apr. 2016, accepted 2 May 2016

Tammet H. 2016: Identification of boundaries between size classes of atmospheric aerosol particles. *Boreal Env. Res.* 21: 363–371.

A set of few non-overlapping size classes is a widely used structure for presentation of atmospheric aerosol measurements. The boundaries between size class intervals are usually decided on the basis of visual analysis of distribution diagrams without explicit quantitative argumentation. I propose a simple and transparent method of moving neighbor correlation for choosing the boundaries between the size class intervals. The method is in a straight line based on the general principle of classification: two items picked from the same class should be similar and two items picked from different classes should be as different as possible. A measure of similarity is the correlation coefficient between the values of the distribution function at two diameters. A correlation coefficient does not depend on constant factors and is the same in the case of the particle volume and number distributions. The method is illustrated with examples based on measurements of atmospheric aerosols at Hyytiälä during the years 2008–2010.

Aerosol measurement is an essential method for obtaining information about processes that control the meteorological phenomena and induce changes in Earth's climate (Seinfeld and Pandis 2012). Particle size spectrometers are powerful tools for atmospheric aerosol research. The output of a particle size spectrometer is a large number of channel signals corresponding to different values of particle diameter D_p. The range of particle sizes in atmospheric air is wide. Thus a logarithmic scale is preferred and the measurements are presented as values of the distribution function $n(D_p) = dN/d\log D_p$ on a logarithmically uniformly divided grid of diameters D_{pi}, $i = 1, \ldots, m$, where $D_{pi+1}/D_{pi} = c$ is a constant. Another interpretation of the measurement record is a set of concentrations N_i in narrow size sections $D_{pi}, \ldots, D_{pi+1}$. The number of sections can be large as, for example, scanning particle sizers (SMPS) manufactured by TSI allow for choosing up to 64 sections per size decade (Price *et al*. 2014). The immediate output of a particle size spectrometer is a source for deriving characteristics that could be basically interpreted and used in solving of practical problems like assessment of air quality.

A popular way is to consider the particle size distribution a composition of a few component distributions and interpret the measurements in terms of parameters of the components. The components may be defined as overlapping or non-overlapping functions on the size scale. The simplest component system is a set of non-overlapping particle size classes where the full size range is split into several subranges called the class intervals, e.g. 3–50 nm, 50–700 nm, and 700–10 000 nm. Thereafter, the classes may

Editor in charge of this article: Veli-Matti Kerminen

be divided into several subclasses according to particle size and physico-chemical properties as explained by Bergman *et al*. (2012). Typically, the size class intervals are declared without explicit argumentation. However, the choice of their boundaries is always somewhat guided by peculiarities of a statistical behavior of particles of different sizes. A well-known system of overlapping components is the multi-modal model (Whitby *et al*. 1991, Hussein *et al*. 2005), where the components are log-normal distributions. Every log-normal component is characterized by its own concentration, mean and standard deviation of $\log D_p$, and the values of these parameters depend on the aerosol type (Jaenicke 1993, John 2011). The location of a peak in the diagram of statistical distribution is called the mode and thus the term "mode" has been adapted for the log-normal aerosol components. The components do not have distinct size limits and the multi-modal model remains still an intuitive guide when choosing the boundaries for non-overlapping classes.

A popular statistical method is correlation analysis, which is often applied when explaining the effect of meteorological conditions and atmospheric trace gases on the formation and development of atmospheric aerosols (Wehner and Wiedensohler 2003). Correlation coefficients between channel signals of an aerosol spectrometer or narrow size fractions of particles make up a large correlation matrix, which is not easy to immediately comprehend. Sophisticated mathematical methods may be used to compress most of the useful information from the matrix into a limited number of parameters, which may simplify the identification of aerosol particle size classes. The most popular method is the principal component analysis. The components are established so that the first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. Original principal components are not associated with definite intervals of particle size and have no simple intuitive interpretation in terms of particle size classes. However, after the second-rate components are dropped, the reduced subset of a few essential principal components can be transformed so that every component distribution has a dominating maximum at certain size of particles (Hõrrak *et al*. 2000, Chan and Mozurkewich 2007). The results of principal component analysis were successfully used as an aid for intuitive classification of atmospheric ions into non-overlapping mobility and size classes by Hõrrak *et al*. (2000).

The aim of the present study was to develop a simple and transparent statistical method, which avoids sophisticated mathematics and supports intuitive allocation of boundaries between class intervals. The method is illustrated and tested using real measurements of atmospheric aerosols.

Data

Real data are necessary for illustrating and testing the method of identifying the particle size classes. Our choice was the open-access data set Hyytiala08_10aerosol (Kulmala and Tammet 2014, 2015) as a tool for exploring the structure and dynamics of atmospheric aerosol size distributions. This data set includes 21 682 hours of routine measurements during three years (2008–2010) and covers a wide size range. The distribution of particles according to their diameter was measured by using a Differential Mobility Particle Sizers (DMPS) with two Hauke-type DMAs (Aalto *et al*. 2001) and Aerodynamic Particle Sizer (APS) TSI model 3321. The measurements with different instruments were merged in accordance with suggestions by Virkkula *et al*. (2011). The DMPS and APS data were overlapping in the border region of 0.53–0.98 μm, which allowed for the verification of the quality of the measurements and smooth transition between the two instruments using weighted averages described by Tammet and Kulmala (2014). The data set is composed of values of the distribution function $dN/d\log D_p$ on a logarithmically-uniform grid of diameters between 2.94 nm and 14.3 μm and presented as an Excel-readable table. The table contains 60 columns of the distribution function $dN/d\log D_p$ and 30 columns of complementary variables. A detailed description of the data structure and application hints are available in a document appended to the data table in the data set by Kulmala and Tammet (2015). The description includes sample dia-

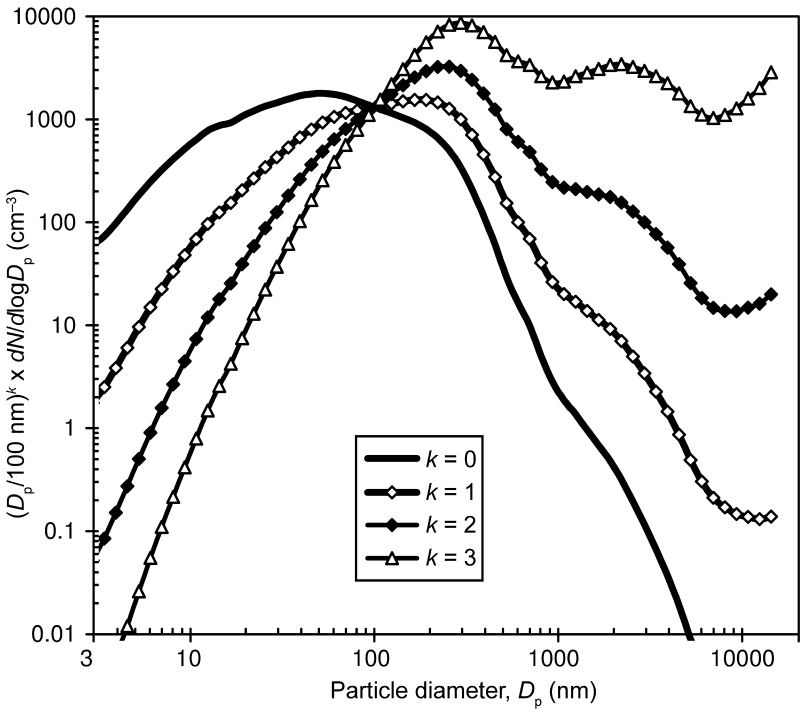

Fig. 1. Mean size distributions $(D_p/100\ nm)^k \times dN/d\log D_p$ of atmospheric aerosol particles recorded in Hyytiälä during 2008–2010. Markers are placed according to 1/16-decade sections.

grams, which illustrate the data and may provoke ideas in atmospheric aerosol research. The mean particle size distribution for all 21 682 hours is illustrated in Fig. 1.

The measurements from Hyytiälä are characteristic of the European boreal forest climate zone. Asmi *et al*. (2011) compared the size distributions of atmospheric aerosol particles measured during 2008–2009 at 24 monitoring stations located in different regions of Europe. The overall concentration of particles varied greatly from site to site, but the shapes of the size distributions at most of the remote ground-level stations incl. Hyytiälä were pretty similar. We believe that the Hyytiälä data well represents the atmospheric aerosol particle size distribution in northern Europe.

Method of moving neighbor correlation

Usually, the boundaries between particle size classes are allocated intuitively when analyzing diagrams of the measured distribution functions. A well-known obstruction stems from the different shapes of diagrams drawn for particle number distributions and for volume or mass distributions. An example in Fig. 2 illustrates the particle size distribution functions for number concentration and for volume concentration in case of a collection of measurements picked from the Hyytiälä data set and containing 75 hours during April midday's. We see two maxima on the curve of the particle number distribution and two maxima on the curve of the particle volume distribution. However, the locations of the maxima on these two curves are completely different. The distribution of the number concentration has maxima at about 9 and 50 nm, while the maxima of the volume concentration are located at about 300 and 2500 nm. A visual analysis of these curves does not provide unambiguous conclusions about the structure of the particle size distribution.

We propose a simple method for identifying the class boundaries, where the analysis of number concentrations and volume concentrations lead to exactly the same diagrams and same conclusions. The method is based on the general principle of classification: two items picked from the same class should be similar as much as pos-

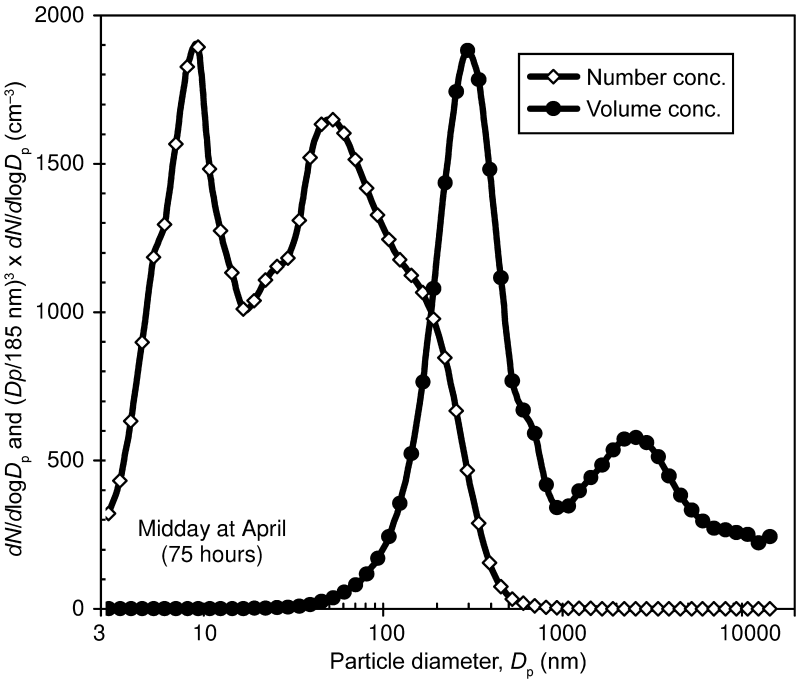

Fig. 2. Size distribution of atmospheric aerosol particles at April midday's, which is a period of intensive events of new particle formation. Measurements are picked from the data set of measurements in Hyytiälä. Curve $(D_p/185\ \text{nm})^3\ dN/d\log D_p$ is proportional to the particle volume distribution. Markers are depicted according to 1/16-decade sections.

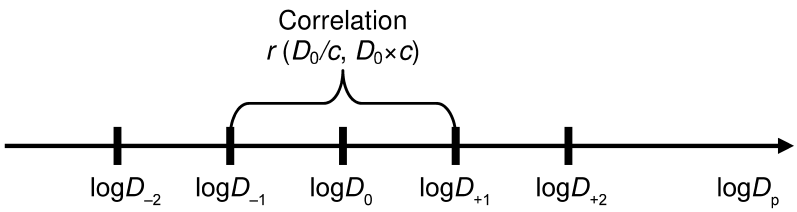

Fig. 3. Argument of moving neighbor correlation D_0 and the closest neighbor diameters.

sible and two items picked from different classes should be different as much as possible. A natural measure of a similarity between variations of particle distribution at two different sizes is the correlation coefficient between the values of the distribution function. The correlation coefficient is independent of constant factors and do not depend on the presentation of arguments, which may be section concentrations or values of distribution function for number, surface or volume of particles.

The Hyytiälä data are presented with values of particle number distribution function at 16 sections per diameter decade. On the logarithmic scale, the diameters of symmetric neighbors of a section with a central diameter of D_0 are $D_{-k} = D_0/c^k$ and $D_{+k} = D_0 \times c^k$, where c is the ratio of successive diameters D_{pi+1}/D_{pi}. The index $k = 1$ corresponds to the closest neighbors and indexes 2, 3, …, to more distant neighbors (Fig. 3). In case of the Hyytiälä data, $c = 10^{1/16} \approx 1.16$ and $c^2 \approx 1.33$. We considered a large set of measurements and calculated moving neighbor correlation coefficient $r_k(D_0)$ between the values of distribution function $n(D_{-k})$ and $n(D_{+k})$ as a function of the central diameter D_0 (Fig. 3) According to the general principle of classification, the correlation should be high if D_0 appears somewhere inside a particle size class interval, and low or negative when D_0 is located just on a boundary between the two size class intervals. The sample neighbor correlation diagram (Fig. 4) is drawn for the same collection of measurements, which was used when compiling the distribution diagrams in Fig. 2. Some visual properties of the curves in Fig. 4 may be a result of imperfections in the instruments. Due to the random noise, the neighbor correlation never reaches exactly 1

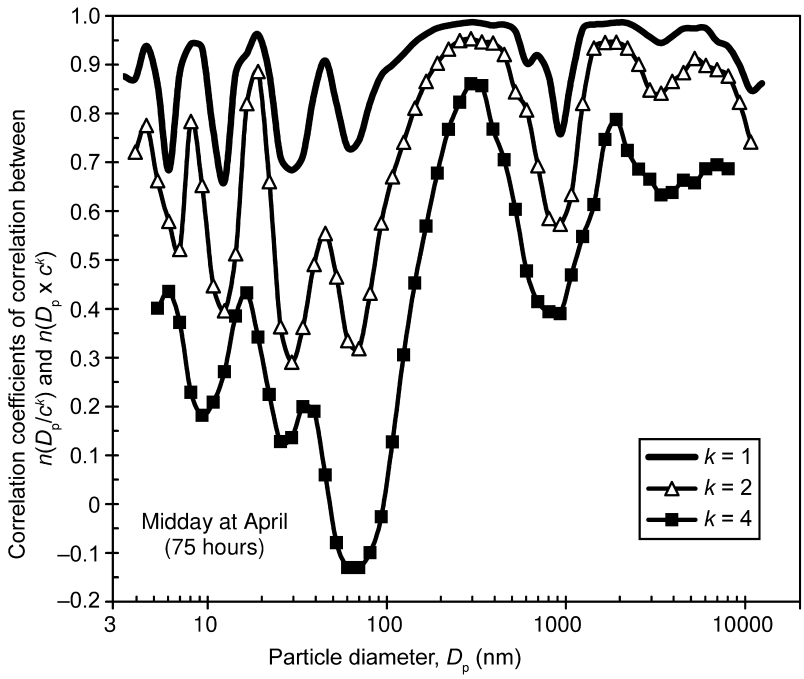

Fig. 4. Moving neighbor correlations for size distribution of atmospheric aerosol particles at $D_p/10^{k/16}$ and $D_p \times 10^{k/16}$ in collection of the same 75 measurements as in Fig. 2. Curves for number distribution and volume coincide exactly. Values of factor c^k are $c^1 \approx 1.16$, $c^2 \approx 1.33$, and $c^4 \approx 1.78$.

even inside an ideal homogeneous size class. The size resolution of a real size spectrometer is limited and the transfer functions of neighbor channels partly overlap. This effect appends some apparent positive correlation, which can turn to be notable near the boundaries between physically-independent or negatively-correlated size classes.

The main findings when comparing the curves in Figs. 4 and 2 were:

— the correlation thresholds (Fig. 4) were much more expressive than the minima in the size distribution curves (Fig. 2),
— the size distribution diagram shows maxima and minima in the distribution function, whereas the correlation diagram immediately indicates the boundaries between internally-correlated size classes,
— the threshold at about 900 nm was visible in both figures, but the expressive threshold at about 65 nm in Fig. 4 could not be identified in Fig. 2.

The choice $k = 1$ at 16 sections per decade ($c^k \approx 1.16$) provides a good size resolution and fair sensitivity. The choice $k = 4$ ($c^k \approx 1.78$) provides a good sensitivity and low size resolution, and can be recommended only in case of very noisy measurements. In a typical situation and 16 sections per decade, the choice $k = 1$ seems to be appropriate. If necessary, the size resolution can be improved by including the intermediate diameters between the points of the main grid. The neighbor correlation at the intermediate diameter $D_0 \times c^{1/2}$ can be calculated as the correlation coefficient between the values of $n(D_0)$ and $n(D_0 \times c)$.

Figures 2 and 4 represent a special period of intensive events of new particle formation, which involves an extra diverse structure of curves in Fig. 4 below 70 nm. A more substantial analysis of this specific situation does not belong to the tasks of the present study, so investigating this further will remain a challenge for future research.

Discussion of neighbor correlation using example of measurements from Hyytiälä

Typical size distributions of atmospheric aerosol particles are similar to the long-term average dis-

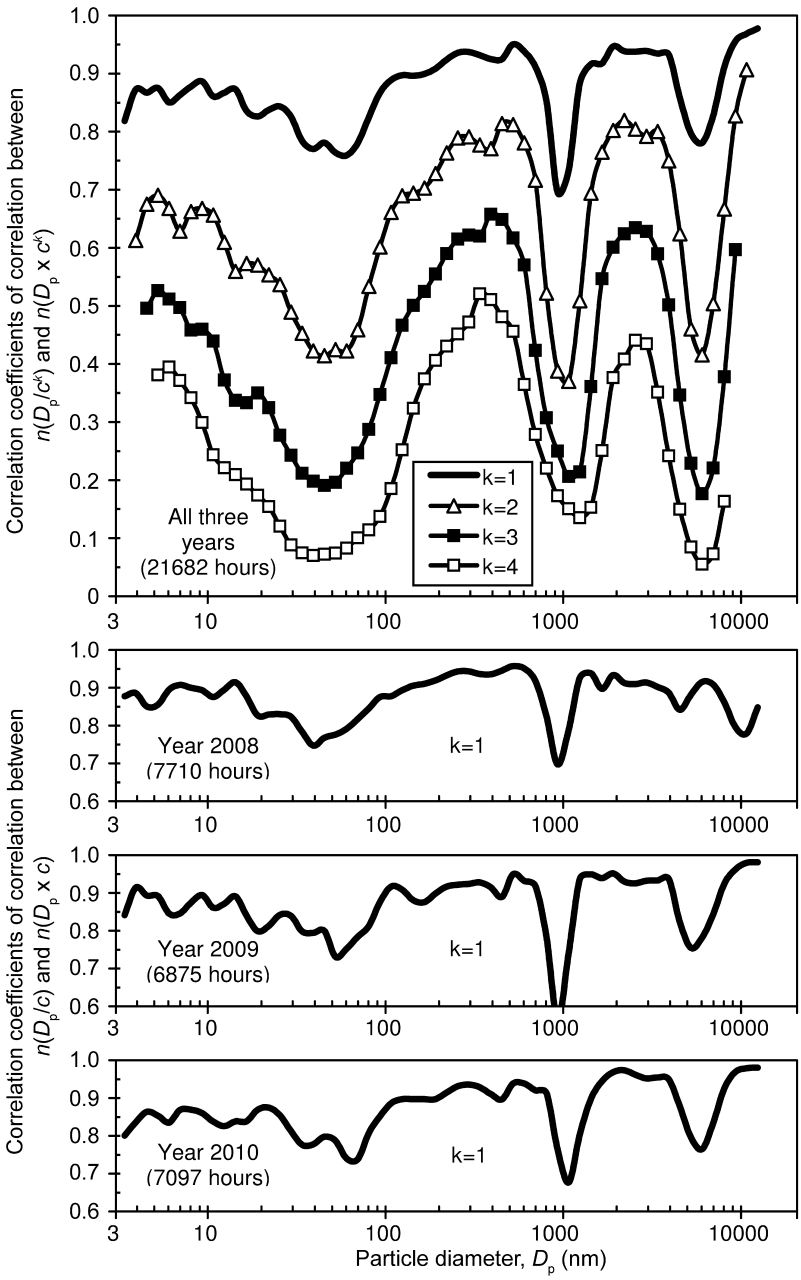

Fig. 5. Moving neighbor correlations for particle size distribution at $D_p/10^{k/16}$ and $D_p \times 10^{k/16}$ according to measurements in Hyytiälä during 2008–2010.

tribution shown in Fig. 1. The neighbor correlation diagrams calculated for the three individual years 2008, 2009 and 2010 differ only slightly from the diagram corresponding to the full three-year data (Fig. 5). The most stable element in all the diagrams is an expressive threshold at about 900 nm, which can be appreciated as a definite boundary and starting point when assembling a system of non-overlapping size classes.

The boundary at about 900 nm is located close to the border between the size ranges of two different instruments: fine particles up to a diameter of 980 nm were measured using DMPS and large particles beginning from diameter of 530 nm by means of APS. Different instruments may bring about different measurement errors. This fact raises some caution when looking for interpretations of particle classes. We still

believe that the threshold of about 900 nm in the case of the Hyytiälä data cannot be explained by measurement errors. Errors caused by switching between two instruments usually have a systematic character. A systematic error means that a measurement differs from the true value by a constant factor. However, systematic multiplication of one variable with a constant can not affect correlations and distort the results of a neighbor correlation analysis.

The lower universal boundary is located near 60 nm. The diagram of a neighbor correlation in the subrange of particles smaller than 60 nm is a bit fuzzy and does not have deep and stable minima. This region is affected by new particle formation events, which tend to generate particle size distributions similar to the curves in Fig. 2. The correlation diagram (Fig. 4) suggests that the smallest particles could be segregated into narrow sub-classes with fuzzy boundaries. Unfortunately, the structure of such a set of subclasses would not be stable.

The upper universal boundary appears at about 6 μm during the years 2009 and 2010. The diagrams for these two years are pretty similar, but the year 2008 turned out to be different in the range of large particles. In 2008, the particles with diameters of around 6–7 μm were well correlated. The upper boundary of micrometer-range particles was shifted toward the upper limit of the data set and might exceed 10 μm. Additionally, the class of micrometer particles is not homogeneous, and the minimum of about 4.5 μm in the diagram suggests an opportunity to discuss splitting the micrometer particles into two subclasses.

The structure of particle size distribution individually for four seasons and for daytime and nighttime were compared with each other (Fig. 6). The main conclusion was that the correlation threshold at about 1 μm was steadily present in all eight time periods. The region below 50 nm was still fuzzy. A remarkable peculiarity was a clear difference between the daytime and nighttime curves, as well as the uniform shape of the daytime curves during the seasons other than winter. The three daytime curves from March to August showed a steep decrease in the neighbor correlation near the lower size limit of the data set. This allows one to expect a hypothetic class boundary below 3 nm depending on new-particle-formation events. The peculiarities in the left wing of the correlation diagrams show the necessity to continue the study involving the measurements of air ions, clusters, and finest nanometer particles.

The problematic upper boundary in a micrometer size range was distinctly present at about 5 μm during autumn. It moved to 6–7 μm in winter, to 9–11 μm in spring and reached the upper size range of the data set in summer. This boundary was a little higher during daytime than during nighttime. The weight of actual warm season in one-year data is subject to climate fluctuations. This probably explains why the diagram for the year 2008 differed from the diagrams for the years 2009 and 2010 (Fig. 5).

Conclusions

Segregating aerosol particles into non-overlapping size classes can be based directly on the general principle of classification: two items picked from the same class should be similar and two items picked from different classes should be different as much as possible. A natural measure of similarity is the correlation coefficient between values of the distribution function. If the size distribution is given on a set of narrow size sections, the identification of boundaries between size class intervals can be based on the correlation between the sections. A correlation coefficient is independent of constant factors and does not depend on the presentation of arguments, which may be section concentrations or values of distribution functions for number, surface or volume of particles. The independence on constant factors ensues that the method is unresponsive for switching between different instruments, which can be used for covering a wide range of particle diameters.

A diagram of moving neighbor correlation is a convenient graphical tool, which assists the identification of natural boundaries between the intervals of particle size classes. Correlation thresholds are graphically much more expressive when compared with stationary points of distribution curves. A correlation diagram indicates immediately the boundaries between internally strongly correlated classes.

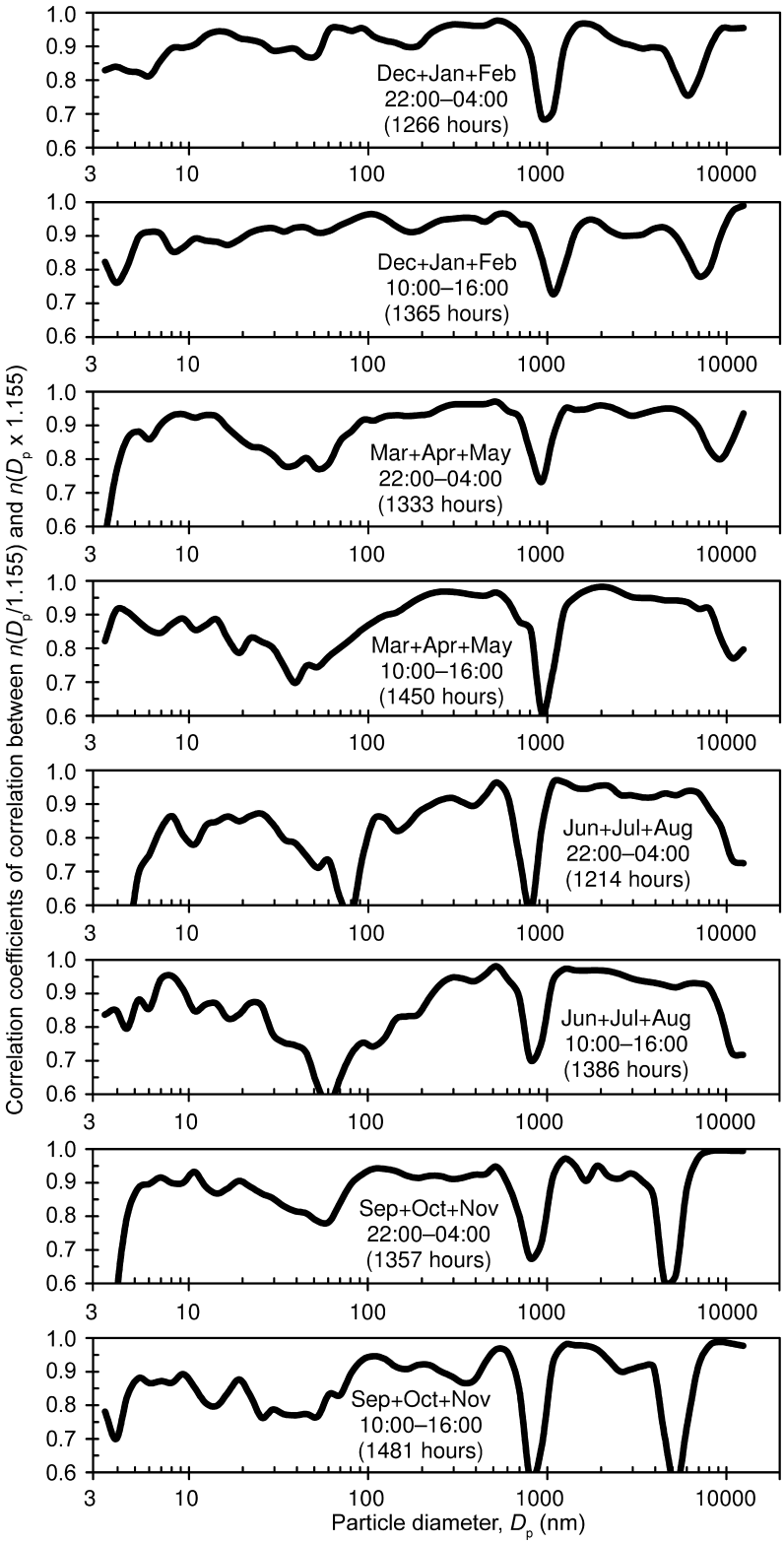

Fig. 6. Moving neighbor correlations for particle size distribution at $D_p/1.16$ and $D_p \times 1.16$ for selected time periods according to measurements in Hyytiälä during 2008–2010.

The natural correlation boundaries between particle class intervals depend somewhat on the season and time of day. There, the boundaries are partially smoothed in a long-term statistics. It was possible to identify three distinct size thresholds at 50–80 nm, 0.8–1 μm, and 5–10 μm in the three-year statistics of measurements from Hyytiälä. The threshold at about 0.9 μm was the most distinct and stable. It was clearly evident during all seasons and times of day, and it had a low variation in time. The threshold at about 70 nm was gentle and clearly presented during summer only. A threshold above 5 μm was always present but its location was not stable. A corresponding minimum in the neighbor correlation curve was deep and located at 5 μm in autumn, shifting to 10 μm and even larger diameters during summer.

The microphysical and chemical background of aerosol particle classifications remain out of scope of the present study. The main statistical size thresholds turn up not far from the particle subrange boundaries 50 and 700 nm accepted in the sectional aerosol microphysics module SALSA (Bergman *et al*. 2012), which can be recommended as a source of knowledge about chemistry and microphysics of size-classified particles.

Acknowledgment: Prof. Markku Kulmala is gratefully acknowledged for scientific support and access to measurements made at SMEAR II, Station for Measuring Forest Ecosystem–Atmosphere Relations in Hyytiälä, Finland.

References

Aalto P., Hämeri K., Becker E., Weber R., Salm J., Mäkelä J.M., Hoell C., O'Dowd C.D., Karlssons H., Hanssons H., Väkevä M., Koponen I.K., Buzorius G. & Kulmala M. 2001. Physical characterization of aerosol particles during nucleation events. *Tellus* 53B: 344–358.

Asmi A., Wiedensohler A., Laj P., Fjaeraa A.-M., Sellegri K., Birmili W., Weingartner E., Baltensperger U., Zdimal V., Zikova N., Putaud J.-P., Marinoni A., Tunved P., Hansson H.-C., Fiebig M., Kivekäs N., Lihavainen H., Asmi E., Ulevicius V., Aalto P.P., Swietlicki E., Kristensson A., Mihalopoulos N., Kalivitis N., Kalapov I., Kiss G., de Leeuw G., Henzing B., Harrison R.M., Beddows D., O'Dowd C., Jennings S.G., Flentje H., Weinhold K., Meinhardt F., Ries L. & Kulmala M. 2011. Number size distributions and seasonality of submicron particles in Europe 2008–2009. *Atmos. Chem. Phys*. 11: 5505–5538.

Bergman T., Kerminen V.-M., Korhonen H., Lehtinen K.J., Makkonen R., Arola A., Mielonen T., Romakkaniemi S., Kulmala M. & Kokkola H. 2012. Evaluation of the sectional aerosol microphysics module SALSA implementation in ECHAM5-HAM aerosol-climate model. *Geosci. Model Dev*. 5: 845–868.

Chan T.W. & Mozurkewich M. 2007. Simplified representation of atmospheric aerosol size distributions using absolute principal component analysis. *Atmos. Chem. Phys*. 7: 875–886.

Hõrrak U., Salm J. & Tammet H. 2000. Statistical characterization of air ion mobility spectra at Tahkuse Observatory: Classification of air ions. *J. Geophys. Res*. 105: 9291–9302.

Hussein T., Dal Maso M., Petäjä T., Koponen I.K., Paatero P., Aalto P.P., Hämeri K., & Kulmala M. 2005. Evaluation of an automatic algorithm for fitting the particle number size distributions. *Boreal Env. Res*. 10: 337–355.

Jaenicke R. 1993. Tropospheric aerosols. In: Hobbs P.V. (ed.), *Aerosol–cloud–climate interactions*, Acadenuc Press, San Diego, CA, pp. 1–31.

John W. 2011. Size distribution characteristics of aerosols. In: Kulkarni P., Baron P.A. & Willeke K. (eds.), *Aerosol measurement*, John Wiley & Sons, Hoboken, NJ, pp. 41–54.

Kulmala M. & Tammet H. 2015. *Hyytiala08_10aerosol*. Univ. Tartu DataDoi Repository, http://dx.doi.org/10.15155/repo-3.

Price H.D., Stahlmecke B., Arthur R., Kaminski H., Lindermann J., Däuber E., Asbach C., Kuhlbusch T.A.J., BéruBé K.A. & Jones T.P. 2014. Comparison of instruments for particle number size distribution measurements in air quality monitoring. *J. Aerosol Sci*. 76: 48–55.

Seinfeld J.H & Pandis S.N. 2012. *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons, Hoboken NJ.

Tammet H. & Kulmala M. 2014. Empiric equations of coagulation sink of fine nanoparticles on background aerosol optimized for boreal zone. *Boreal Env. Res*. 19: 115–126.

Virkkula A., Backman J., Aalto P.P., Hulkkonen M., Riuttanen L., Nieminen T., Dal Maso M., Sogacheva, L. De Leeuw G. & Kulmala M. 2011. Seasonal cycle, size dependencies, and source analyses of aerosol optical properties at the SMEAR II measurement station in Hyytiälä, Finland. *Atmos. Chem. Phys*. 11: 4445–4468.

Wehner B. & and Wiedensohler A. 2003. Long term measurements of submicrometer urban aerosols: statistical analysis for correlations with meteorological conditions and trace gases. *Atmos. Chem. Phys*. 3: 867–879.

Whitby E.R., McMurry P.H., Binkowski F. & Shankar U. 1991. *Modal aerosol dynamics modeling*. Report for contract No. 68-01-7365, U.S. Env. Prot. Agency, Research Triangle Park, NC.